Page 13, importance sampling formula

$$\mathbb{E}_{x\sim p(x)}[H(x)] = \int_x p(x)H(x)dx = \int_x q(x)\frac{p(x)}{q(x)}H(x)dx = \mathbb{E}_{x\sim q(x)}[\frac{p(x)}{q(x)}H(x)]$$

Page 14, Kullback-Leibler divergence

$$KL(p_1(x)\|p_2(x)) = \mathbb{E}_{x\sim p_1(x)} \log\frac{p_1(x)}{p_2(x)} = \mathbb{E}_{x\sim p_1(x)}[\log p_1(x)] - \mathbb{E}_{x\sim p_1(x)}[\log p_2(x)]$$

Page 14, text snippet

The first term in KL is called entropy and doesn't depend on $p_2(x)$, so, could...

Combining both formulas, we can get the following iterative algorithm, which starts with $q_0(x) = p(x)$, and on every step improves approximation of $p(x)H(x)$ with update

$$q_{i+1}(x) = \underset{q_{i+1}(x)}{\arg\min} - \mathbb{E}_{x\sim q_i(x)}\frac{p(x)}{q_i(x)}H(x)\log q_{i+1}(x)$$

Page 14, policy update

$$\pi_{i+1}(a|s) = \underset{\pi_{i+1}}{\arg\min} - \mathbb{E}_{z\sim\pi_i(a|s)}[R(z)\geq\psi_i]\log\pi_{i+1}(a|s)$$